A Competitiveness-driven and Secure Incentive Mechanism for Competitive Organizations Data Sharing: A Contract Theoretic Approach

Bingyi Guo*, Xiaofang Deng†, Quansheng Guan‡ and Jie Tian*
*Sch.of Information Science and Engineering, Shandong Normal University, Jinan, P.R. China
†Sch. of Information and Communication, Guilin University of Electronic Technology, Guilin, P.R. China
‡Sch. of Information and Electronic Engineering, South China University of Technology, Guangzhou, P.R. China

***Abstract*—In the era of big data and artificial intelligence, data sharing is desirable for vigorous development of data-driven intelligent services. Although data sharing is supported to a certain extent by current mechanisms and technologies, organizations especially with potential competitive relationships might refuse to share their data. One reason is that data holders worry that data sharing improves competitors' competitiveness. The other reason is that data sharing suffers huge privacy security risk. To address these problems, in this paper, the concept of competitiveness is introduced as a data sharing transaction driving force to eliminate the competitiveness worry of data holders while differential privacy is adopt to protect their privacy. As there is an information asymmetry between data sharers and data demanders, a contract theoretic approach is proposed to motivate data holders to share data with privacy protection, which is expected to achieve a target of win-win and data sharing security. By designing optimal contracts, the data demander can decide rationally how to pay the data holders given the privacy parameter. Moreover, data holders can choose the contract that maximize their utilities. Numerical results substantiate the effectiveness of the the proposed scheme.**

Keywords*—Data sharing, incentive mechanism, competitiveness, privacy protection, contract theory

I. Introduction

With the rapid development of data processing technology, data-driven services such as recommendation services, speech recognition and image recognition emerge vigorously, which changes the style of daily life. These intelligent services are provided by organizations by processing a sufficient amount of high quality data. However, not all the data-driven service providers can possess the same wealth of data in the large corporations, such as Google, Facebook, Microsoft and Amazon [1]. Actually, sharing data among multiple organizations is an efficient approach to address the data availability issue, and improve significantly data-driven services. The sharing of scientific data can advance the progress of scientific research, and provide a redundant backup for valuable data set [2]. The sharing of financial data can help detect fraud and other illegal activities by searching links between transfers [3]. Data from hospitals can help predict flu outbreaks, and then improve the response to epidemics [1].

Although data sharing is highly desired for vigorous development of data-driven services, organizations especially with competitive relationships might refuse to share data [4]. One reason is that data holders worry that data sharing improves competitors' competitiveness. With the help of data sharing, competitors improve the user experience of their services based on the acquired data processing while data holders reduce business due to the loss of data. This difficult situation hinders data holders' motivation of data sharing. For instance, data sharing and collaboration is critical to the research for new medicines [5]. However, medicine sectors are often driven by profit and discovering the next super drug so that they have no passion to share data [6]. The other reason is that data sharing suffers huge privacy security risks [3]. When data is disclosed or used in data-driven services with data sharing, individual's privacy will be inevitably compromised [7], [8]. Therefore, the privacy threats posed by data sharing are concerned increasingly.

In this sense, in order to motivate data sharing to improve current intelligent services, incentive mechanisms are required for organizations to eliminate the competitiveness worry and guarantee data sharing security. Specially, we need to develop mechanisms that ensure getting strictly better service for organizations who provides data while protecting their privacy. Although data sharing is supported by current mechanisms and technologies, most of previous works are focused on data sharing among participants with cooperative relationships. How to motivate competitive organizations to share data is largely ignored [9], [10]. Moreover, the privacy protection of data sharing among competitive organizations is still a problem to be addressed. Many current mechanisms are not credible because the privacy protection is carried in the third party rather than data sources [11]. Even if these third parties are considered as trustful, there is still privacy disclosure risk.

Therefore, in the data sharing among competitive organizations, it is a challenge to eliminate the competitiveness worry and guarantee privacy security. In this paper, we jointly consider competitiveness motivation and privacy protection by introducing the concept of competitiveness as a data sharing transaction driving force and employing privacy preservation at data holders with differential privacy. As the exact privacy parameters is unknown for data demanders, a contract theoretic approach is proposed to address the information asymmetry. By designing optimal contracts, the data demander can decide rationally how to pay the data holders given the privacy parameter. Moreover, data holders can optimize their utilities by choosing a best contract. Numerical results substantiate the effectiveness of the the proposed scheme.

The rest of the paper is organized as follows. The system model is presented in Section II. The contract-theoretic model is formulated in Section III. The theoretic and discrete optimal contract design are discussed in Section IV and Section V,

978-1-5386-4870-4/18/$31.00 ©2018 IEEE

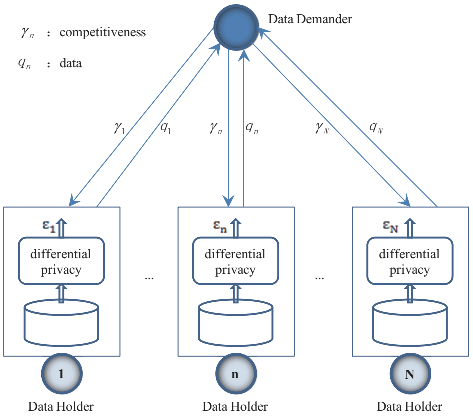


Fig. 1. A data sharing network.

respectively. Simulation results are discussed in Section VI. Finally, we conclude this study in Section VII.

II. Model Description

A. Competitiveness Model

In data sharing among competitive organizations, data holders worry that competitors acquire a amount of data and then improve their competitiveness through technical progress and even technology leapfrogging. As data holders can not benefit from the data sharing, they are in a disadvantageous position in commercial competition, reducing the passion of data sharing. However, if data holders obtain the technical progress and competitiveness in data sharing, there will be a possible win-win result for data holders and data demanders. Therefore, as defined in Definition 1, the concept of competitiveness is an important incentive factor for data sharing.

Definition 1 (Competitiveness)**.** *Competitiveness is the ability to improve the quality of services such as user experience, which is usually obtained by processing data.*

- *Competitiveness of a data set can be measured by a factor γ called competitive factor.*
- *A data set with high competitive factor brings a data holder more profit.*

In this paper, competitiveness is used as a transaction driving force to motivate data sharing. In this sense, data holders share their data to the data demander, and obtain the ability of technical progress from the data demander, i.e., competitiveness expressed in terms of competitive factor. Thus, data holders eliminate the competitiveness worry by obtaining the competitiveness.

B. Differential Privacy

To provide the privacy security of data holders during data sharing, differential privacy is conducted before sharing the data [12]. Differential privacy is a strong and rigorous standard for privacy protection. Some well-known software systems such as Google Chrome [13] and Apple iOS [14] have applied this new technology to protect users' privacy. Differential privacy guarantees that no third party can infer individual-level information with high confidence based on the released results.

As a gold standard of privacy definition, differential privacy also is a rigorous mathematic definition. Suppose that D and $D^{'}$ are neighboring databases, which are differing on at most one record. M is a random function and P_M is the any output of M. For D and $D^{'}$, a privacy mechanism M gives ε-differential privacy if it satisfies

$$\Pr[M(D) \in S_M] \leq \exp(\varepsilon) \times \Pr[M(D^{'}) \in S_M],$$

where $S_M \subseteq P_M$. The privacy parameter ε is the privacy budget. The higher value of ε corresponds to the lower privacy protection. The privacy parameter is usually very small. In the data sharing network as shown in Fig. 1, before sharing the data to the data demander, data holders take ε-differential privacy to preserve their privacy, which provides a secure data sharing mode.

C. System Model

Consider a data sharing network in Fig. 1. The data sharing network consists of N data holders and one data demander, which have competitive relationships. On the requirement to improve the quality of data-driven services, the data demander collects data from N data holders. Once handing over its data, the data holder may suffer the loss of data control, potential improvement of the data demander's competitiveness and a potential loss in privacy. To protect the privacy security, data holders adopt differential privacy before sharing data. A privacy parameter $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$ is used to describe the protection level of the privacy. According to differential privacy, a smaller ε means the data holder takes better privacy protection. The privacy parameter is defined by the data holder, which is unknown to the data demander. As the competitiveness is introduced, when a data holder shares data to the data demander, it will receive competitiveness as compensation for the data loss from the data demander. In this data sharing network, each data demander and data holder make decisions to maximize their utilities.

From the perspective of the data demander, it collects data from data holders to achieve service improvement. For the data collected q, The gain from data q, $G(q)$ is expressed as

$$G(q) = \omega \log(1 + q), \tag{1}$$

where ω is a positive parameter meaning the weight of the data to the data demander. Before sharing data, data holders take privacy protection action by differential privacy. For the collected data q from the data holder with the privacy parameter ε, a small ε may cause a decrease in data utility because the small ε means high privacy protection and less details of the shared data. Once receiving the data q from the data holder with the privacy parameter ε, the data demander should pay competitiveness to the data holder. Therefore, the overall utility of the data demander by obtaining data from a data holder with the privacy parameter ε can be defined as

$$U_D = \omega \log\left(1 + \alpha\varepsilon q\right) - \gamma, \tag{2}$$

where α is a positive parameter.

From the perspective of a data holder, it takes $\varepsilon-$differential privacy to protect privacy before sharing data

q. Although the data holder suffers potential privacy loss, it will receive the competitiveness as compensation for the loss of data. Therefore, the utility to a data holder with privacy parameter ε can be defined as

$$u = \gamma - c\varepsilon q, \tag{3}$$

where c is a cost parameter.

In the data sharing network, the exact privacy parameters is unknown to the data demander because the differential privacy is taken privately by the data holders. In other words, the privacy parameter ε is a random variable to the data demander. In this paper, we assume that $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$ is drawn independently and identically for different data holders. The data demander only knows the probability density function $f(\varepsilon)$ and cumulative distributed function $F(\varepsilon)$. Therefore, there is an information asymmetry between the data demander and data holders.

III. Contract-Theoretic Formulation

To resolve the conflicts between the data demander and data holders in the presence of asymmetric information, a contract-theoretic approach is proposed in this section. Each data holder's privacy parameter ε is not observable to the data demander. To achieve data sharing, the data demander offers a menu of contracts $\{(\gamma, q)\}$ to each data holder. After receiving the menu of contracts, each data holder will choose one contract (γ, q) that maximizes its utility. Thus, according to the contract accepted, the data holder provides data q to the date demander, and in return, the data demander should pay competitiveness in terms of competitive factor γ to the data holder. According to the revelation principle [8], it is sufficient for the data demander to consider the contracts that ensure each data holder to truthfully choose the contract designed for its privacy parameter. Therefore, the contract is designed as a pair of functions $\{(\gamma(\varepsilon), q(\varepsilon))\}$, where the contract $(\gamma(\varepsilon), q(\varepsilon))$ is designated for data holder with privacy parameter ε. In this sense, upon choosing the contract $(\gamma(\varepsilon), q(\varepsilon))$, the utility of a data holder with privacy parameter ε can be expressed as

$$u_\varepsilon(\gamma(\varepsilon), q(\varepsilon)) = \gamma(\varepsilon) - c\varepsilon q(\varepsilon). \tag{4}$$

To ensure that data holders will accept the contracts designated for them rather than choosing other contracts or refusing any contract, the contracts must be incentive feasible. To be a feasible contract, $\{(\gamma(\varepsilon), q(\varepsilon))\}$ needs to satisfy both the incentive compatibility constraints and the individual rationality constraints.

Definition 2 (Incentive Compatibility (IC))**.** *A menu of contracts* $\{(\gamma(\varepsilon), q(\varepsilon)), \varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]\}$ *satisfies IC if the data holder with privacy parameter* ε *prefers to accept the contract* $(\gamma(\varepsilon), q(\varepsilon))$ *rather than other contracts, i.e.,*

$$u_\varepsilon(\gamma(\varepsilon), q(\varepsilon)) \geq u_\varepsilon(\gamma(\hat{\varepsilon}), q(\hat{\varepsilon})), \forall \varepsilon, \hat{\varepsilon} \in [\underline{\varepsilon}, \overline{\varepsilon}]. \tag{5}$$

Definition 3 (Individual Rationality (IR))**.** *A menu of contracts* $\{(\gamma(\varepsilon), q(\varepsilon)), \varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]\}$ *satisfies IR if each data holder has a non-negative utility by accepting the contract for its privacy parameter* ε*, i.e.,*

$$u_\varepsilon(\gamma(\varepsilon), q(\varepsilon)) \geq 0, \forall \varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]. \tag{6}$$

In the contract theoretic data sharing model, since the data demander offers contracts to data holders without knowing the data holders' privacy parameters, the utility of the data demander is evaluated in expected terms. The data demander's objective is to find an optimal menu of contracts which maximizes the expected utility. Therefore, the data demander's objective can be formulated as

$$\begin{gathered} \max N \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} U_D(q(\varepsilon), \gamma(\varepsilon)) f(\varepsilon) d\varepsilon \\ \text{subject to (5) and (6),} \end{gathered} \tag{7}$$

where the function $U_D(q(\varepsilon), \gamma(\varepsilon))$ is defined as

$$U_D(q(\varepsilon), \gamma(\varepsilon)) = \omega \log(1 + \alpha \varepsilon q(\varepsilon)) - \gamma(\varepsilon). \tag{8}$$

IV. Optimal Contract Design

A. Simplifying the Optimization Problem

Proposition 1. *A menu of contracts* $\{(\gamma(\varepsilon), q(\varepsilon)), \varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]\}$ *satisfies IC constraint if and only if*

$$q^{'}(\varepsilon) \leq 0, \tag{9}$$

$$\gamma^{'}(\varepsilon) - c\varepsilon q^{'}(\varepsilon) = 0, \tag{10}$$

where $q^{'}(\varepsilon) = \dfrac{dq(\varepsilon)}{d\varepsilon}$ *and* $\gamma^{'}(\varepsilon) = \dfrac{d\gamma(\varepsilon)}{d\varepsilon}$.

Proof: According to the IC definition in (5), for any $\varepsilon, \hat{\varepsilon} \in [\underline{\varepsilon}, \overline{\varepsilon}]$, we have

$$\gamma(\varepsilon) - c\varepsilon q(\varepsilon) \geq \gamma(\hat{\varepsilon}) - c\varepsilon q(\hat{\varepsilon}), \tag{11}$$

$$\gamma(\hat{\varepsilon}) - c\hat{\varepsilon} q(\hat{\varepsilon}) \geq \gamma(\varepsilon) - c\hat{\varepsilon} q(\varepsilon). \tag{12}$$

Adding the above two inequalities, we have

$$(\varepsilon - \hat{\varepsilon}) c (q(\hat{\varepsilon}) - q(\varepsilon)) \geq 0. \tag{13}$$

As $c > 0$, the above inequality means that $q(\varepsilon)$ is a non-increasing function of ε. Therefore, it can be concluded that

$$q^{'}(\varepsilon) \leq 0. \tag{14}$$

Given ε, (11) implies that the utility function of the data holder with privacy parameter ε, $u_\varepsilon(\gamma(\hat{\varepsilon}), q(\hat{\varepsilon})) = \gamma(\hat{\varepsilon}) - c\varepsilon q(\hat{\varepsilon})$ reaches its maximum at $\hat{\varepsilon} = \varepsilon$. Therefore, we have

$$\frac{\partial u_\varepsilon(\gamma(\hat{\varepsilon}), q(\hat{\varepsilon}))}{\partial \hat{\varepsilon}}\Big|_{\hat{\varepsilon}=\varepsilon} = \frac{d\gamma(\varepsilon)}{d\varepsilon} - \varepsilon c \frac{dq(\varepsilon)}{d\varepsilon} = 0. \tag{15}$$

Next, it will be proved that conditions in (9) and (10) are also sufficient conditions for the IC constraint. According to (10) , we have

$$\gamma(\varepsilon) - \gamma(\hat{\varepsilon}) = \int_{\hat{\varepsilon}}^{\varepsilon} \gamma^{'}(\tau) d\tau = \varepsilon c q(\varepsilon) - \hat{\varepsilon} c q(\hat{\varepsilon}) - \int_{\hat{\varepsilon}}^{\varepsilon} c q(\tau) d\tau. \tag{16}$$

After some manipulations of (16), according to (9), we have

$$\begin{aligned} \gamma(\varepsilon) - c\varepsilon q(\varepsilon) &= \gamma(\hat{\varepsilon}) - \varepsilon c q(\hat{\varepsilon}) + \int_{\hat{\varepsilon}}^{\varepsilon} [c q(\hat{\varepsilon}) - c q(\tau)] d\tau \\ &\geq \gamma(\hat{\varepsilon}) - \varepsilon c q(\hat{\varepsilon}), \end{aligned} \tag{17}$$

which follows the definition of IC constraint. ■

Proposition 2. *Suppose a menu of contracts* $\{(\gamma(\varepsilon), q(\varepsilon)), \varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]\}$ *satisfies the IC constraint. Then, the menu of contracts satisfies the IR constraint if and only if*

$$\gamma(\overline{\varepsilon}) - c\varepsilon q(\overline{\varepsilon}) \geq 0. \tag{18}$$

Proof: Since the menu of contracts satisfies the IC constraint, according to (10), we have

$$u_\varepsilon^{'} = \frac{du_\varepsilon(\gamma(\varepsilon), q(\varepsilon))}{d\varepsilon} = \gamma^{'}(\varepsilon) - cq(\varepsilon) - c\varepsilon q^{'}(\varepsilon) = -cq(\varepsilon). \tag{19}$$

Since $c > 0$ and $q(\varepsilon) \geq 0$, we have $-cq(\varepsilon) \leq 0$, which indicates that $u_\varepsilon(\gamma(\varepsilon), q(\varepsilon))$ is a non-increasing function of ε. Therefore, $\overline{\varepsilon}$ is the privacy parameter minimizing $u(\varepsilon)$, i.e.,

$$\overline{\varepsilon} \in \min_{\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]} u_\varepsilon(\gamma(\varepsilon), q(\varepsilon)). \tag{20}$$

In this sense, the IR constraint is thus equivalent to $u_{\overline{\varepsilon}}(\gamma(\overline{\varepsilon}), q(\overline{\varepsilon})) \geq 0$, i.e., $\gamma(\overline{\varepsilon}) - c\varepsilon q(\overline{\varepsilon}) \geq 0$. ■

Theorem 1. *For the optimal solution, the IR constraint for the privacy parameter* $\overline{\varepsilon}$ *is binding at the optimum, i.e.,*

$$u_{\overline{\varepsilon}}^{*} = 0. \tag{21}$$

Proof: Suppose $u_{\overline{\varepsilon}}^{*} > 0$, then the data demander could reduce $u_{\overline{\varepsilon}}^{*}$ by a small amount while keeping $q^*(\varepsilon)$ unchanged. As a result, the data demander's utility is increased, which contradicts with the optimality of $u_{\overline{\varepsilon}}^{*}$. ■

Based on the simplifications, the optimization problem of the data demander is rewritten as

$$\begin{aligned} &\max \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} U_D(q(\varepsilon), \gamma(\varepsilon)) f(\varepsilon) d\varepsilon \\ &= \max \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \Big(\omega \log\left(1 + \alpha\varepsilon q(\varepsilon)\right) - u_\varepsilon(\gamma(\varepsilon), q(\varepsilon)) \\ &\qquad - \varepsilon c q(\varepsilon) \Big) f(\varepsilon) d\varepsilon \\ &\qquad \text{subject to } (9), (10) \text{ and } (18). \end{aligned} \tag{22}$$

B. Optimal Control-based Approach

The Pontryagin's maximum principle is used for solving the optimization problem of the data demander in (22) to obtain the optimal data function $q^*(\varepsilon)$. Let $q(\varepsilon)$ be the control variable, $u_\varepsilon(\gamma(\varepsilon), q(\varepsilon))$ be the control variable and ε be the time variable. Let $x(\varepsilon) = u_\varepsilon(\gamma(\varepsilon), q(\varepsilon))$. The Hamiltonian of the problem is expressed as

$$\begin{aligned} &H(\mathbf{x}(\varepsilon), q(\varepsilon), \lambda(\varepsilon), \varepsilon) \\ &= \Big(\omega \log\left(1 + \alpha\varepsilon q(\varepsilon)\right) - x(\varepsilon) - \varepsilon c q(\varepsilon) \Big) f(\varepsilon) - \lambda(\varepsilon) c q(\varepsilon). \end{aligned} \tag{23}$$

According to the Pontryagin minimum principle, the necessary conditions of the optimal control and states are as follows.

$$x^{'}(\varepsilon) = \frac{\partial H(\mathbf{x}^*(\varepsilon), q^*(\varepsilon), \lambda^*(\varepsilon), \varepsilon)}{\partial \lambda(\varepsilon)} = f(\varepsilon) q^*(\varepsilon). \tag{24}$$

$$\lambda^{'}(\varepsilon) = -\frac{\partial H(\mathbf{x}^*(\varepsilon), q^*(\varepsilon), \lambda^*(\varepsilon), \varepsilon)}{\partial x(\varepsilon)} = f(\varepsilon). \tag{25}$$

$$H(\mathbf{x}^*(\varepsilon), q^*(\varepsilon), \lambda^*(\varepsilon), \varepsilon) \geq H(\mathbf{x}^*(\varepsilon), q(\varepsilon), \lambda^*(\varepsilon), \varepsilon). \tag{26}$$

```
Algorithm 1 Optimal Contract Implementation.
  1. Optimal Contract Design:
       The data demander calculates the optimal contract
       $\{\gamma^*(\varepsilon), q^*(\varepsilon)\}$ according to (29) and (31).
  2. Contract Publish:
       The data demander publishs the contract $\{\gamma^*(\varepsilon), q^*(\varepsilon)\}$
       to all data holders.
  3. Contract Selection:
       Each data holder decides whether to accept the contract
       or not according to its own utility.
  4. Data Sharing:
       Once accepting the contract, the data holder shares the
       data and receives the competitiveness according to the
       contract.
```

From above conditions, we have

$$\lambda^*(\varepsilon) = F(\varepsilon). \tag{27}$$

As having the $\boldsymbol{\lambda}^*(\varepsilon)$, the optimal data function can be derived via taking the first-order derivative of Hamiltonian to $q(\varepsilon)$. We get

$$\begin{aligned} \frac{\partial H(\mathbf{x}^*(\varepsilon), q(\varepsilon), \lambda^*(\varepsilon), \varepsilon)}{\partial q(\varepsilon)} &= f(\varepsilon)\left(\frac{\omega\alpha\varepsilon}{1 + \alpha\varepsilon q(\varepsilon)} - \varepsilon c \right) \\ &\quad - cF(\varepsilon) = 0. \end{aligned} \tag{28}$$

From (28), we have the optimal $q^*(\varepsilon)$ as

$$q^*(\varepsilon) = \frac{\omega}{\varepsilon c + cF(\varepsilon)/f(\varepsilon)} - \frac{1}{\alpha\varepsilon}. \tag{29}$$

With the optimal data function $q^*(\varepsilon)$, we can obtain the optimal utility function of data holders $u_\varepsilon^*(\gamma(\varepsilon), q(\varepsilon))$ and competitiveness function $\gamma^*(\varepsilon)$ according to (19) and (4) as

$$u_\varepsilon^*(\gamma(\varepsilon), q(\varepsilon)) = \int_{\overline{\varepsilon}}^{\varepsilon} -cq^*(\tau) d\tau, \tag{30}$$

$$\gamma^*(\varepsilon) = c\varepsilon q(\varepsilon) + \int_{\overline{\varepsilon}}^{\varepsilon} -cq^*(\tau) d\tau. \tag{31}$$

C. Optimal Contract Implementation

According to the proposed mechanism, the procedure of optimal contract implementation is as shown in Algorithm 1.

V. Discrete Optimal Contract Design

In practice, it is difficult for the data demander to get the distribution of the privacy parameter. In this section, by making the privacy parameter discrete, we try to design more practical optimal contracts. We quantize the set of privacy parameters $\Theta = [\underline{\varepsilon}, \overline{\varepsilon}]$ with a factor K such that the privacy parameters are a discrete set of K privacy parameters, i.e., $\Theta = \{\delta_1, \delta_2, \cdots, \delta_K\}$. Without loss of generality, it can be assumed that $\delta_1 < \delta_2 < \cdots < \delta_K$. The quantization process is considered to be uniform with equidistant values, i.e., $\delta_k = \underline{\varepsilon} + (k-1)\sigma$ where $\sigma = \dfrac{\overline{\varepsilon} - \underline{\varepsilon}}{K}$.

If K is large enough, the privacy parameters of data holders are almost equal to a δ_k in Θ. The objective of the data demander is to maximize its expected utility by designing an incentive compatible and individually rational optimal contract $(\gamma(\delta_k), q(\delta_k))$ (for simplicity, we will now refer it as (γ_k, q_k)) $\forall \delta_k \in \Theta$. Therefore, the objective function in (7) can be rewritten in the discrete form as

$$\begin{aligned} &\max \sum_{k=1}^{K} \left(\omega \log\left(1+\alpha\delta_k q_k\right) - \gamma_k\right) \\ &s.t. \quad \gamma_k - c\delta_k q_k \geq \gamma_j - c\delta_k q_j, \forall \delta_k, \delta_j \in \Theta. \\ &\qquad\quad \gamma_k - c\delta_k q_k \geq 0, \forall \delta_k \in \Theta. \end{aligned} \tag{32}$$

Theorem 2. *For the optimal solution, the individual rationality constraint for the highest privacy parameter is binding, i.e.,* $\gamma_K - c\delta_K q_K = 0$.

Proof: According to Theorem 1, for the optimal solution, the IR constraint for the privacy parameter $\bar{\varepsilon}$ is binding at the optimum. In the discrete form, we have $\gamma_K - c\delta_K q_K = 0$. ■

Theorem 3. *For the optimal solution,* $q_1 \geq q_2 \geq \cdots \geq q_K \geq 0$ *and all the upward adjacent ICs are binding, i.e.,*

$$\gamma_k - c\delta_k q_k = \gamma_{k+1} - c\delta_k q_{k+1}, \forall k \leq K-1. \tag{33}$$

Proof: According to Proposition 1, the data function is a non-increasing function of privacy parameter, we have

$$q_1 \geq q_2 \geq \cdots \geq q_K \geq 0. \tag{34}$$

1) Proof of sufficiency: Suppose the adjacent upward ICs are binding, then we can rewrite (33) as follows

$$\gamma_k - \gamma_{k+1} = c\delta_k(q_k - q_{k+1}), \forall k \leq K-1 \tag{35}$$

For some j such that $j > k$, then using (34) and (35), we have

$$\begin{aligned} &\gamma_k - \gamma_j = c\delta_k(q_k - q_{k+1}) + c\delta_{k+1}(q_{k+1} - q_{k+2}) \\ &+ \cdots + c\delta_{j-1}(q_{j-1} - q_j) \\ &\geq c\delta_k(q_k - q_{k+1}) + c\delta_k(q_{k+1} - q_{k+2}) + \cdots \\ &+ c\delta_k(q_{j-1} - q_j) = c\delta_k(q_k - q_j). \end{aligned} \tag{36}$$

Then, for some j such that $j < k$, then using (34) and (35), similarly we have

$$\gamma_j - \gamma_k \leq c\delta_k(q_j - q_k). \tag{37}$$

Combining (36) and (37), we have the IC constraint expression

$$\gamma_k - c\delta_k q_k \geq \gamma_j - c\delta_k q_j, \forall k \neq j. \tag{38}$$

Hence all the ICs are satisfied when adjacent upward ICs are binding.

2) Proof of necessity: Suppose there are one or more upward adjacent ICs, such as δ_k, which are not binding for optimal solutions. With IR for δ_{k+1}, we get

$$\gamma_k - c\delta_k q_k > \gamma_{k+1} - c\delta_k q_{k+1} \geq \gamma_{k+1} - c\delta_{k+1} q_{k+1} \geq 0. \tag{39}$$

If we reduce all γ_j, $\forall j < k$ with equal data, it will no change for any IRs and the existing relation between adjacent upward ICs. We iteratively repeat the process (from the highest privacy parameter for which adjacent upward IC is inactive) till all the upward ICs are binding. During this process, we have only reduced the competitiveness to bind all the upward adjacent ICs. This in turn satisfies all the other ICs from the sufficiency conditions. Hence, we find a better contract and the original contract cannot be optimal, which is a contradiction. ■

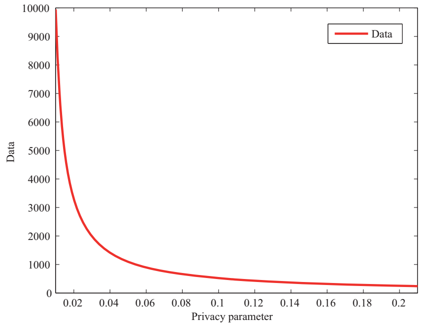


Fig. 2. Data function vs. privacy parameters.

Thus, the function in (32) can be expressed as

$$\max \sum_{k=1}^{K} \left(\omega \log\left(1+\alpha\delta_k q_k\right) - \gamma_k\right), \tag{40}$$

where

$$\gamma_k = \begin{cases} \gamma_{k+1} + c\delta_k(q_k - q_{k+1}), & \forall k \leq K-1, \\ c\delta_K q_K, & k = K. \end{cases} \tag{41}$$

Taking the derivative of the optimization function to q_k and equating it to zero, we have the optimal data function

$$q_k = \frac{\omega}{kc\delta_k - (k-1)c\delta_{k-1}} - \frac{1}{\alpha\delta_k}. \tag{42}$$

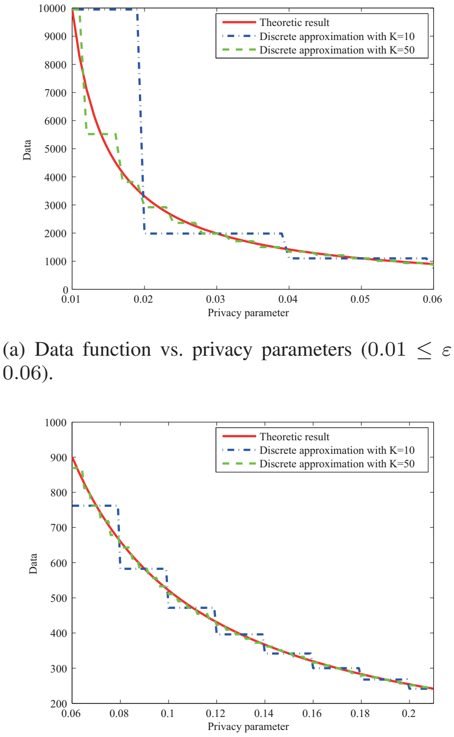
The optimal competitiveness function can be correspondingly obtained by combining (41) and (42).

VI. Numerical Results and Discussions

In this section, numerical simulations are conducted to study the performance of the proposed contract-based data sharing incentive mechanism. For ease of illustration, we consider a simple data sharing network with one data demander and 5 data holders. We assume the privacy parameters of data holders are distributed within $[0.01, 0.21]$ uniformly. Data weight $\omega = 100$. The data processing complexity factor $\alpha = 2$ and the data sharing cost $c = 1$.

The data function performance of the theoretic optimal contract is presented in Fig. 2, which shows that the data function in optimal contract decreases with the increase of the privacy parameter. The reason is that a large privacy parameter value means a large possibility of privacy disclosure according to the definition of differential privacy so that the data holders are reluctant to share data, resulting in the decrease of the data in optimal contract.

The comparison of the discrete optimal contract function with the theoretic results is presented in Fig. 3 and Fig. 4. Fig. 3 shows the performance of the discrete optimal data function, where the data function in the discrete optimal contract design is close to theoretic results. For more clear expression, the details of the comparison are drawn in Fig. 3(a) and Fig. 3(b). As K increases, the discrete data function is getting closer to theoretic results. Moreover, the two curves almost coincide when $K = 50$, which verifies that the data function in discrete



(a) Data function vs. privacy parameters ($0.01 \leq \varepsilon \leq 0.06$).

(b) Data function vs. privacy parameters ($0.06 \leq \varepsilon \leq 0.21$).

Fig. 3. Data function vs. privacy parameters.

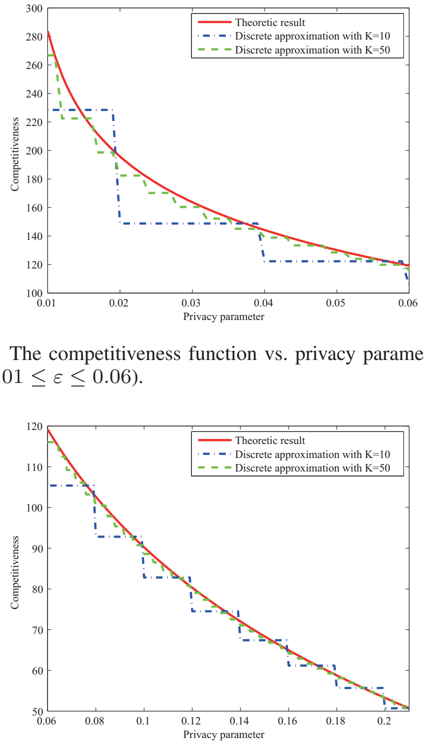
optimal contract design is a good approximation with a large K. The performance of the discrete optimal competitiveness function is depicted in Fig. 4, which shows that the competitiveness function in the discrete optimal contract design is close to theoretic results. Therefore, the discrete optimal contract design approximates the theoretic results well.

VII. Conclusions and Future Work

In this paper, an incentive mechanism based on contract theory has been proposed to address the competitiveness worry problem and privacy security problem in data sharing among organizations with potential competitive relationships. By introducing competitiveness into data sharing as a motivation factor, data demanders and data holders are encouraged to participate in the data sharing. Moreover, privacy is preserved by employing differential privacy. As there is an information asymmetry among data sharing participants, the incentive mechanism is formulated as a contract theoretic approach to achieve a target of win-win and data sharing security. By designing an optimal contract, data holders and the data demander can maximize their utilities. Numerical results show the effectiveness of the proposed scheme.

Acknowledgment

This work is supported by Guangxi Provincial Education Department Foundation (No.2017KY0202), Guangxi Cognitive Radio and Information Processing Key Laboratory Authorized by China's Ministry of Education Foundation (No.CRKL150110) and 2017 Open Foundation (No.CRKL170203), Guangxi Broadband Wireless Communication and Signal Processing Key Laboratory Foundation (No.GXKL06160114), Shandong Provincial Natural Science Foundation for Young Scholars of China (No.ZR2017QF008) and the Open Research Fund from Shandong Provincial Key Laboratory of Computer Network (No. SDKLCN-2016-02).



(a) The competitiveness function vs. privacy parameters ($0.01 \leq \varepsilon \leq 0.06$).

(b) The competitiveness function vs. privacy parameters ($0.06 \leq \varepsilon \leq 0.21$).

Fig. 4. The competitiveness function vs. privacy parameters.

References

[1] I. Stoica, D. Song, R. A. Popa, D. Patterson, M. W. Mahoney, R. Katz, A. D. Joseph, M. Jordan, J. M. Hellerstein, J. E. Gonzalez, *et al.*, "A berkeley view of systems challenges for ai," *arXiv preprint arXiv:1712.05855*, 2017.

[2] R. A. Poldrack and K. J. Gorgolewski, "Making big data open: data sharing in neuroimaging," *Nature Neuroscience*, vol. 17, no. 11, pp. 1510–1517, 2014.

[3] M. A. Will, R. K. Ko, and S. J. Schlickmann, "Anonymous data sharing between organisations with elliptic curve cryptography," in *Proc. IEEE Trustcom/BigDataSE/ICESS*, pp. 1024–1031, 2017.

[4] F. Skopik, G. Settanni, and R. Fiedler, "A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing," *Computers & Security*, vol. 60, pp. 154–176, 2016.

[5] B. Fabian, T. Ermakova, and P. Junghanns, "Collaborative and secure sharing of healthcare data in multi-clouds," *Information Systems*, vol. 48, pp. 132–150, 2015.

[6] P. Groves, B. Kayyali, D. Knott, and S. V. Kuiken, "The "big data" revolution in healthcare. accelerating value and innovation," *Mckinsey & Company*, pp. 4,13–16, 2016.

[7] M. Dhanalakshmi and V. Balu, "Effective incentive compatible model for privacy preservation of information in secure data sharing and

publishing," *International Journal of Computer Applications*, vol. 96, no. 9, pp. 7–11, 2014.

[8] L. Xu, C. Jiang, Y. Chen, Y. Ren, and K. J. R. Liu, "Privacy or utility in data collection? a contract theoretic approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, pp. 1256–1269, Oct 2015.

[9] X. Wang, L. Ding, Q. Wang, J. Xie, T. Wang, X. Tian, Y. Guan, and X. Wang, "A picture is worth a thousand words: Share your real-time view on the road," *IEEE Trans. on Vehicular Technology*, vol. 66, no. 4, pp. 2902–2914, 2017.

[10] D. Kwak, R. Liu, D. Kim, B. Nath, and L. Iftode, "Seeing is believing: Sharing real-time visual traffic information via vehicular clouds," *IEEE Access*, vol. 4, pp. 3617–3631, 2017.

[11] M. Ali, R. Dhamotharan, E. Khan, S. U. Khan, A. V. Vasilakos, K. Li, and A. Y. Zomaya, "Sedasc: Secure data sharing in clouds," *IEEE Systems Journal*, vol. 11, no. 2, pp. 395–404, 2017.

[12] C. Dwork, "Differential privacy," in *International Colloquium on Automata, Languages, and Programming*, pp. 1–12, 2006.

[13] V. Pihur and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *In Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, p-p. 1054–1067, 2014.

[14] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang, "Privacy loss in apple's implementation of differential privacy on macos 10.12," *arXiv preprint arXiv:1709.02753.*, 2017.